
"Space: The Final Frontier"

Christopher R. Hertel

Alvarri, Inc.

November, 2006

Introduction

Who am I?

- Network Geek
- Storage Geek
- Samba/CIFS Geek
- Author (shameless plug)
- Incurable Idealist

Current Gig: Alvarri, Inc. (a startup)

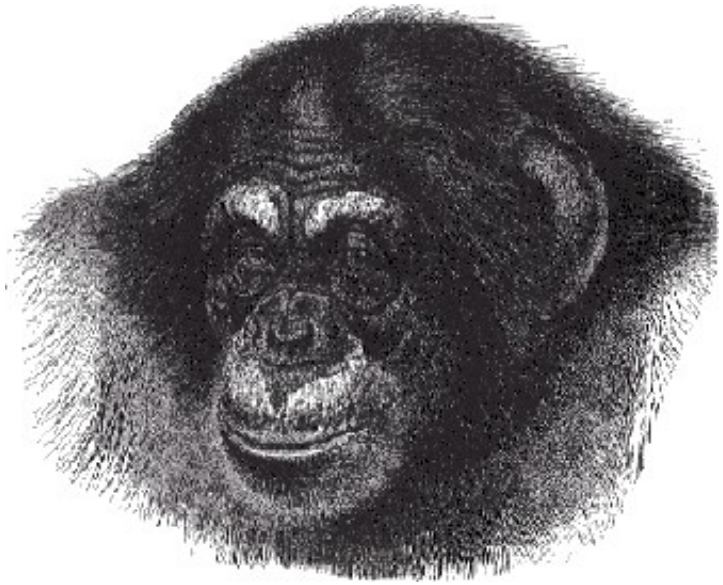
www.alvarri.com

A ruminant mammal (Geekus geekus) with long legs, humped shoulders, and broadly palmated antlers.

Introduction

Who are You?

- Students
- System Administrators
- Network Managers
- Security Geeks
- Coders
- Hackers (per RFC 1392)
- The Morbidly Curious

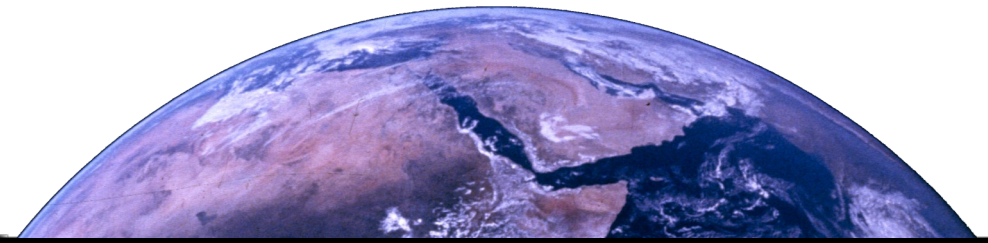


Introduction

Where are we going?

A Tour of Storage Technologies:

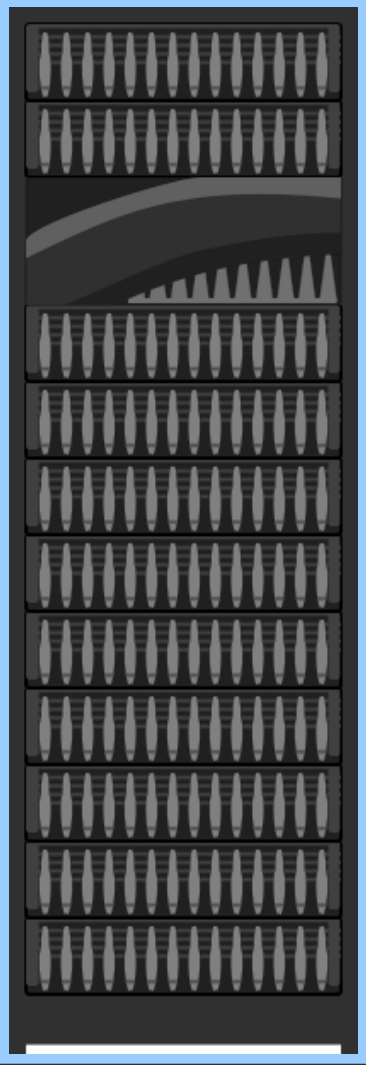
- Disk– 50 Years Young
- SAN – Shared Block Storage
- NAS – Networked File Systems
- Other Things You Will Encounter in your Travels.



A Place for your Stuff

(That's really what disk drives are all about.)

A Place for your Stuff



Sidebar

Disk-o-matic Math

Drive makers measure by 1000, not 1024.

1PB = 1000TB = 909.5 “real” TB
1TB = 1000GB = 931.3 “real” GB
1GB = 1000MB = 953.7 “real” MB
1MB = 1000KB = 976.5 “real” KB
1KB = 1000B

Operating Systems typically use powers of 2 (e.g., $2^{10} = 1024$).
One “real” Petabyte = 2^{50} bytes.

A Place for your Stuff

IBM RAMAC (1956)

Random **A**ccess **M**ethod of **A**ccounting and **C**ontrol

Original Disk Drive:

- Fifty 24″ Platters
- Less Than Five Megabytes (4.4MB)

A Place for your Stuff

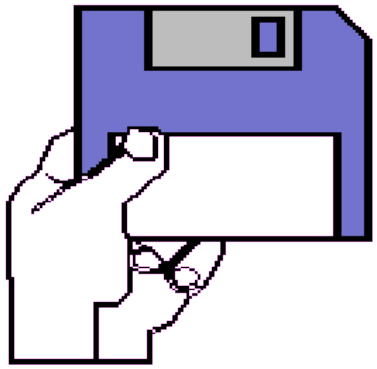
25 years ago: 10MB was *a lot* of disk space.

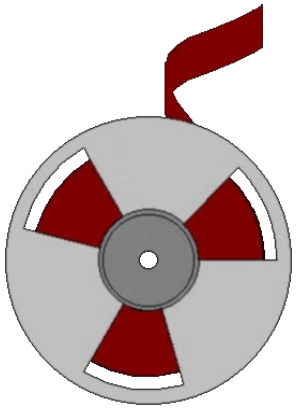
Today: I've got at least 1TB at home.

- 3.5" Drives are $0.20/GB
- Enterprise Storage is measured in Petabytes
- We carry Gigabytes in our pockets

Storage capacity, like computing power, has grown such that we can now have in a handheld what used to require a computer room *and* a team of experts.

A Place for your Stuff





In our increasingly digital world:

- We keep getting more Digital Stuff (data)
- Our Digital Stuff keeps getting bigger (Gigs)
- We worry about keeping our Digital Stuff safe

A Place for your Stuff

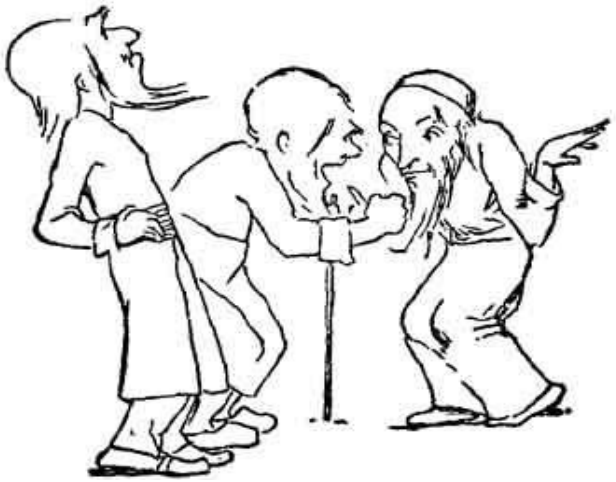
All of that storage...

...scattered all over the home

...scattered all around the office

...scattered all across the Internet

How do we handle it all?



A Place for your Stuff

"The problems that the Lunatic Fringe is working on today are the problems that the mainstream storage industry will face in 5-10 years."

➔ Tom Ruwart,
Storage on the Lunatic Fringe

(He's right, you know.)

Storage on the Lunatic Fringe
http://www.dtc.umn.edu/resources/ruwart.ppt

A Place for your Stuff

Hertel's Corollary: The large-scale storage problems of yesterday afternoon have already become the home office / small office storage problems of early this morning.

Small 1TB JBOD units are available for about $400. (Roughly 700 “real” GB @ RAID5 3+1.)

What's good for the goose...

Benefits of consolidated storage for small-end users:

- Centralized management
- Efficient use of resources
- Data protection (RAID / Backup / Archive)
- Failure isolation

There are problems with centralization, so mix of local and central storage is often the most workable choice.

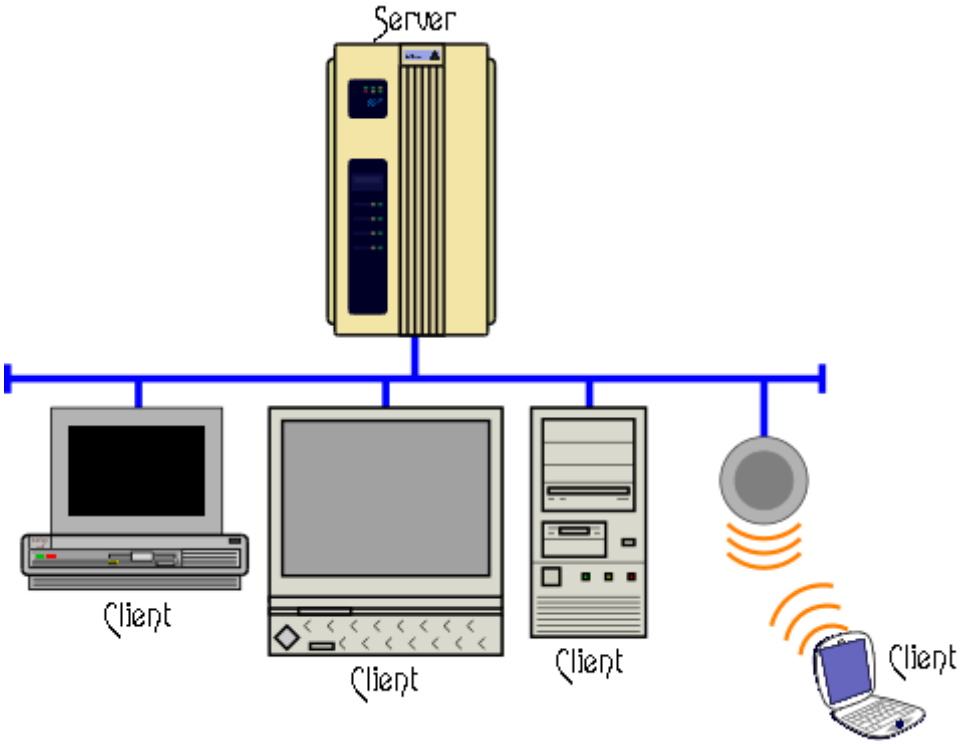

Network Attached
Storage

Familiar NAS Systems:

- IBM (& Microsoft's) SMB/CIFS
- Novell's NetWare
- Apple's Appleshare
- Sun's NFS
- IETF WebDAV

Local file systems on the server are shared with multiple hosts across a LAN or inter-network.

NAS



Typical client/server NAS

- Large server with local disk
- Multiple clients
- Shared access to files & directories

NAS Concerns:

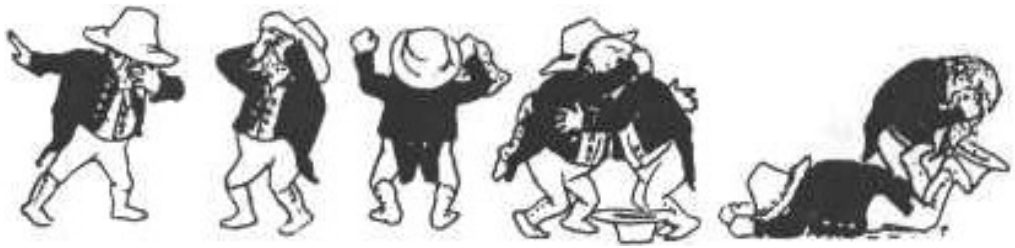
- Authentication, Authorization, & Access Management
- File Locking & Sharing
- Meta-data Semantics

DOS FAT	MacOS	Windows NTFS	Linux/Unix
• System, Hidden, and Archive bits • No UID/GID • 8.3 Format • EOLN: <CR><LF>	• Data and Resource Forks • EOLN: <CR>	• Extended Attributes • File Streams • NT ACLs • EOLN: <CR><LF>	• User, Group, World permission bits • UID/GID • POSIX ACLs • EOLN: <LF>

NAS File Systems are “Vendor Biased”.

Case In Point: CIFS vs. NFS

- For a geek, NFS is easy:
  - Traditionally server-to-server
  - Traditionally geek-to-geek
  - Simple authentication model
- For a user, CIFS is easy:
  - Traditionally user-to-server or peer-to-peer
  - Non-technical user community
  - Specifications and protocol details are hidden



WebDAV

- An extension of HTTP
- Makes the web “read/write”
- Adds only seven new commands
- Messages passed in XML format

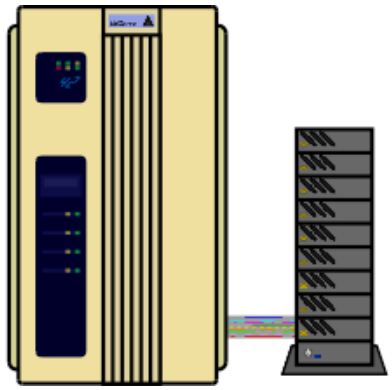
The use of XML
allows great flexibility
... and complexity.

“...as simple as possible, but no simpler.”

This is a picture of my dog.


Storage
Area
Networks

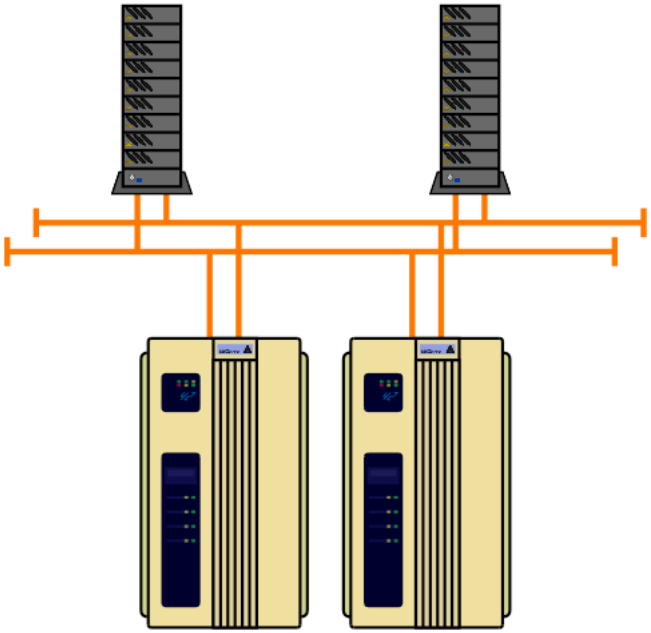
SAN Overview



Precursor: Direct Attached Disk Arrays

- Redundant Array of Inexpensive Disk
- Expandable
- “Virtualizable” (Is that a word?)

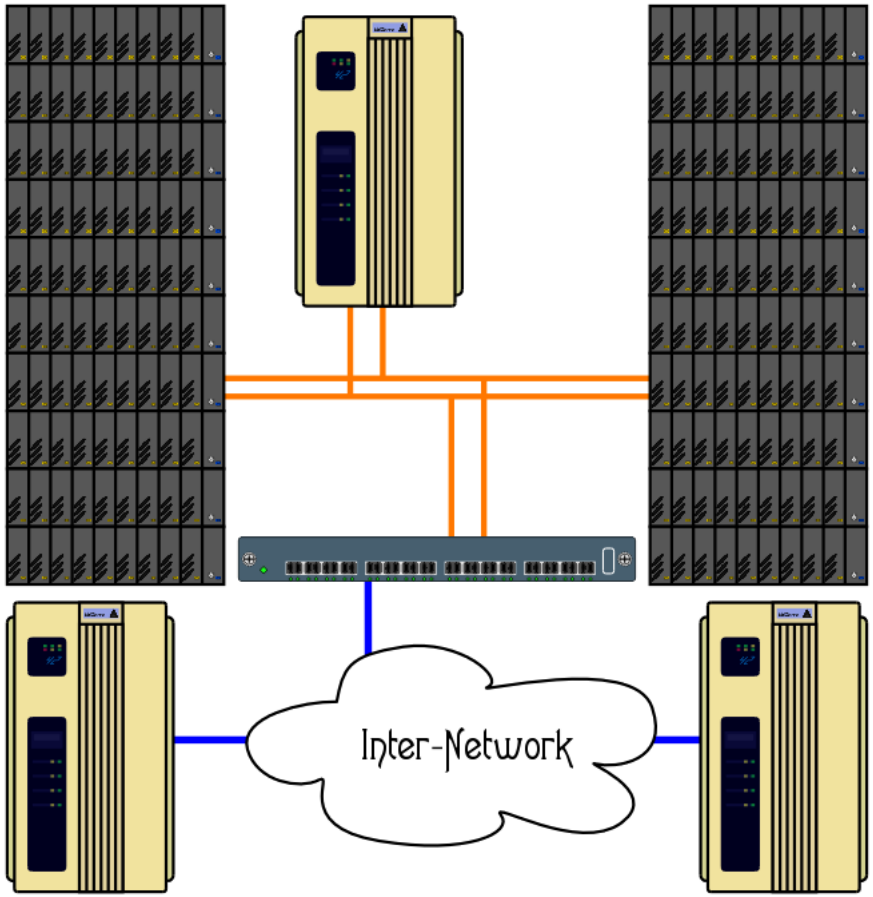
SAN



FibreChannel SANs

- SCSI over Shared/Switched Fiber
- Longer Distances
- 1, 2, 4, and soon 8 & 10 Gbps Speeds
- Redundancy

SAN



iSCSI SANs

- Leverage the IP Network
- Coexist with FibreChannel
- Run on Commodity Network Hardware

SCSI
is the Traditional SAN Protocol

- FibreChannel carries SCSI
- iSCSI is just SCSI PDUs over TCP/IP

The message is the same;
only the transport changes.

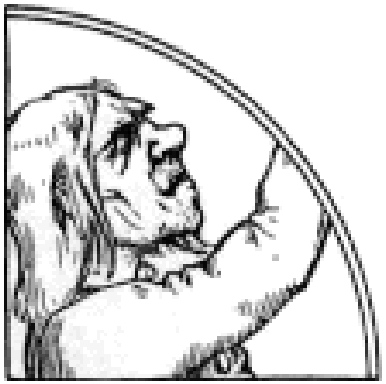
Rivals

- Network Block Dæmon (nbd) for Linux uses TCP/IP as a transport
- AoE (ATA over Ethernet) transports ATA commands over Ethernet frames

SAN vs. NAS

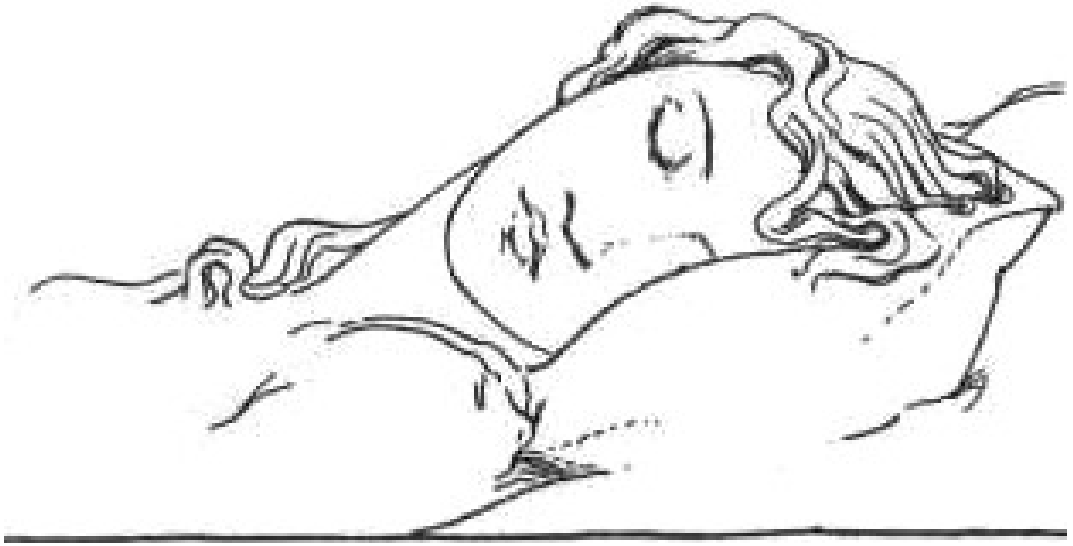
SAN

- Block Storage
- One-to-One Relationship
- Data-center Oriented
- Space is Not Shared



NAS

- File System Storage
- One-to-Many Relationship
- End-User Oriented
- Data Can Be Shared



Other Stuff

MAID: Massive Array of Idle Disks

- Cheap Disks (Commodity ATA)
- Densely Packed
- Mostly Powered Down
- Presented as (virtual) Tape Libraries

Idle drives are spun up from time to time to ensure that they don't get stuck.

More than an interface — SCSI vs. ATA
http://www.seagate.com/docs/pdf/whitepaper/D2c_More_than_Interface_ATA_vs_SCSI_042003.pdf

Other Stuff

ILM: Information Lifecycle Management

- Identify different storage classes
  - high speed vs. low speed
  - high availability vs. high latency
  - expensive vs. cheap
- Monitor data access
- Migrate data (manually/automatically)

For example, migrate from RAID1+0 SCSI drives to RAID5 ATA to Tape.

Other Stuff

Linux: Your Storage Playpen

- Home SAN:
  - AoE and iSCSI
- FUSE: User Mode File System Interface
  - E.g.: SSH, FTP, and BitTorrent clients
- Logical Volume Manager (LVM)
- Software RAID

Other Stuff

Unusual Beyond the Strange

- Cluster File Systems
  - E.g.: Global File System (GFS)
- Distributed File Systems
  - E.g.: Google File System (GFS)
- Object File Systems
  - E.g.: Lustre and UofM T-10 OSD

The End

Slides available at: http://ubiqx.org/presentations/